7 Kant on Trolleys and Autonomous Driving

Abstract: Unlike most interpreters, I do not take it for granted that Kant’s decision would be to divert the trolley. Rather, I submit, the trolley ought not to be turned, given the Kantian difference between narrow and wide duties. It is a narrow duty not to kill innocent human beings, but only a wide duty to save them. Given the latitude of the wide duty (to save the five) and the narrow and negative character of the other duty (not to kill the one), it is morally obligatory to obey the latter, i.e., not to kill the one. Also, because of the latitude of the wide duty, no conflict of duties arises. I shall rebut three objections against this wide/narrow interpretation (as I will call it) and present three additional arguments supporting it. In conclusion, I will discuss the relevance of these findings for the ethical challenges that come along with autonomous vehicles.

New technical developments lead to new ethical problems. The development of autonomous vehicles is a case in point: In view of the advancing automation of road and freight traffic, a decision has to be made as to how autonomous vehicles shall be programmed with regard to possible situations of conflict. For instance, may (or should) such vehicles steer around a group of people in order not to hit them, putting up instead with hitting a smaller group of people or a single person? Part of the theoretical framework to be taken into account in answering this question is the trolley problem. In this paper, I shall take a Kantian perspective on the trolley problem and thus also on the problem of autonomous driving. The final question, then, is: How should, against the background of Kantian ethics, autonomous vehicles be programmed when it comes to trolley situations?

To begin with, I shall spell out the basic features of the trolley problem (1). I will then discuss the question of whether the programmer of autonomous vehicles is to be paralleled with the trolley driver or rather with the so-called bystander (2). Afterwards, I will present a solution to the trolley problem that I think Kant himself would have given – I call it the wide/narrow interpretation (3). I shall then rebut three objections (4) and present three additional arguments supporting that interpretation (5). In conclusion, I will discuss the relevance of these findings for the ethical challenges that come along with autonomous vehicles (6).
1 The Trolley Problem

The trolley problem challenges philosophers to take a stance on very basic and crucial ethical questions, and it can be adapted not only to autonomous driving, but also to quite a number of controversial debates such as just war, torture, or triage. In her famous 1967 paper, Philippa Foot describes the following situation: “[. . .] it may rather be supposed that he [someone] is the driver of a runaway tram which he can only steer from one narrow track on to another; five men are working on one track and one man on the other; anyone on the track he enters is bound to be killed” (Foot 1978 [1967], 23).1 If the driver does not turn the trolley, the five will die; if he does turn the trolley, the five survive but the one will die.2 Two questions arise: First, should the trolley be turned or not? I call this the preceding trolley problem. It is only preceding in the sense that, as a matter of fact, this question typically is given a rather quick and straightforward affirmative answer; according to the prevailing opinion, the trolley may or even must be diverted.3 Therefore, the main question in the literature and so, in this sense, the prevalent trolley problem (as I call it) is this: Why is it permissible to turn the trolley while it is not permissible not to turn it (or to make analogous decisions) in somewhat similar cases?4 For example, it is not permissible to kill one healthy man in order to donate his organs to save some other five, and likewise, it is not permissible to throw a large man down a bridge over the tracks in order to block the trolley with his body before it kills the five (the so-called footbridge case).5

Foot’s solution both to the preceding and the prevalent trolley problem is as follows: The trolley driver has to divert the trolley to save the five. As a rule, in situations in which both negative duties (such as the duty not to kill) and positive duties (such as the duty to help others) apply, negative duties trump positive

1 Foot was not the first to come up with the trolley problem. This is not the place to investigate the historical question of who introduced the trolley problem (or a similar situation) first. In any event, Engisch (1930, 288) and Welzel (1951, 51 f.) already discussed similar cases.
2 In morally relevant respects, the involved persons are equal.
5 Foot contrasts the original trolley case with the case of a judge who has to decide whether he should sacrifice one innocent person to prevent a violent riot or not. She also discusses a number of similar but slightly different cases.
duties; killing, for example, is worse than not helping and thus worse than letting die (rule 1). In the trolley case, however, only negative duties are relevant, so that the driver is faced – Foot argues (1978 [1967], 27) – with a “conflict of negative duties”: If he does nothing, he kills five; if he diverts the course, he kills one; both killings are, generally speaking, prohibited. Since in such a case, when duties of the same kind (such as the negative duty not to kill one and the negative duty not to kill five) are conflicting, the number of victims is to be minimized (rule 2), the driver must divert the trolley. By contrast, a healthy person may not be killed in order to save five by organ donation because the five who would otherwise die of organ failure would merely be left to die, while the healthy person would be killed. In the trolley case, no one would be left to die, but would be killed either way.

As we shall see, this thought has considerable flaws or, in any case, questionable premises. Foot has sparked a tremendous discussion with an enormous number of more or less similar cases (from bridges to bombs, avalanches, trap doors, tunnels, lazy Susans, tractors, and much more) which all drive at a better account of the prevalent trolley problem, i.e., a better answer to the question of why it is permissible to do something in this case (original trolley case) but not in another (transplantation case, large man etc.). Contrary to what usually happens, I shall not primarily deal with those different modifications of the original trolley case and the prevalent trolley problem. Rather, I shall pursue the preceding trolley problem against the background of Kantian ethics, that is, the question of whether it is appropriate to turn the trolley or not (even though the answer to this preceding problem is, of course, related to the solution to the prevalent trolley problem, insofar as the answer to the latter shall provide criteria that also explain why the diversion of the trolley is permissible in the original case). Thus, although the preceding trolley problem is usually relegated to the background of discussion insofar as it seems clear to most that the trolley should be turned, it is precisely this question that will be brought into focus here. The main question then is: Would Kant agree to turn the trolley in the original trolley case?

---

6 Foot does not explicitly speak of rules, but it is obvious that she avails herself of these principles. Thomson (1985, 1396 f.), referring to Foot, articulates the principles (rules).
7 Foot (1978 [1967], 27) does not speak of a duty not to kill at this point but of the “duty to avoid injuring”, but that does not make any difference.
8 For a discussion of the most important cases, cf. Kamm (2016).
2 Driver, Bystander, and Programmer

Before I address the main question, another distinction deserves attention. Obviously, the preceding and the prevalent trolley problem are highly interesting topics. However, my focus in this paper is not primarily on these problems as such, but on their relation to Kant’s ethics and especially to the question of how a Kantian programmer would program fully autonomous vehicles for cases that are similar to the original trolley case. Based on this presupposition, one ought to clarify the Kantian programmer’s point of view by analogy: Does the fact that a certain action is morally permissible or obligatory for the driver of the trolley in the original case really imply that the corresponding programming is permissible or obligatory for the programmer, too? Although the parallelization of programmer and driver seems to be appropriate at first, on closer inspection some doubts arise. For in the context of the trolley problem it is not only the driver that is made a subject of discussion, but also the so-called bystander – and it could be, for instance, that a certain action, which is permissible for one (the bystander), is not permissible for the other (the driver). If this is the case, clarification is called for which role the programmer must fill.

This leads us to Judith Thomson, who introduced the bystander into the discussion. By referring to a situation in which it is not the driver who has to make the decision to turn the trolley, but an uninvolved bystander next to the scene, she aimed at showing that Foot’s solution to the trolley problem is not convincing. Thomson’s argument goes like this: Since the driver of the trolley has started a process that, if he does not intervene, inevitably leads to the death of the five (he started the trolley and steered it up to the point in question), the driver kills the five on the tracks even if he does not do anything. The bystander, on the other hand, finds himself at a short distance from the trolley and has no history with it; he has not set the trolley in motion and thus does not kill the five in case of non-intervention, but only lets them die. However, he would certainly kill the one if he were to throw the switch he is standing next to. According to Thomson, Foot’s rule 1 (negative duties trump positive duties) implies that it is not permissible for the bystander to turn the switch, because by turning the switch he would kill the

---

9 There are different levels of automation when it comes to autonomous vehicles. I will concentrate here on the highest level (full driving automation), where no intervention of the driver is possible and where a driver is not necessary at all.
11 “[. . .] if the driver fails to turn his trolley, he does not merely let the five track workmen die; he drives his trolley into them, and thereby kills them” (Thomson 1985, 1397); “[. . .] if he does not turn the trolley, he does drive his trolley into them, and does thereby kill them” (1398).
one, i.e. the bystander would violate a negative duty which is worse than letting five die. However, according to Thomson’s intuition, it is indeed permissible for the bystander to turn the switch. Consequently, it cannot be true that killing is always worse than letting die (because it is permissible for the bystander to kill a person by turning the switch). Thomson then offers her own answer to the question of why it is permissible to turn the trolley in the driver’s and in the bystander’s case but not in others (the large man case, for instance), but this is of no interest to us here.12

Now the question arises as to whether the role of the programmer of autonomous vehicles is analogous to the role of the trolley driver or to that of the bystander; note, however, that this question is only important if there really is a normative difference between being in the driver’s and the bystander’s position. If the programmer accepts rules similar to Foot’s (negative duties trump positive duties; when duties of the same kind conflict, the number of victims is to be minimized), it follows that, understanding the programmer as a driver, it is permissible for him to turn the trolley; understanding him as a bystander, it is not permissible for him to flip the switch on the basis of those rules. Yet if the Kantian rules (which we still have to determine) are different from Foot’s, it might turn out that the difference between driver and bystander is irrelevant for the programmer.

Admittedly, in a certain respect the programmer is neither driver nor bystander. He is not surprised by a situation in which he is directly involved and in which he has to decide quickly. Rather, the programmer has enough time for his decision.13 And yet these differences are irrelevant, because even in hypothetical trolley situations the crucial question is which carefully considered decision is the right one; psychological pressure or time pressure hardly play any

12 Thomson’s first approach (1985, 1403) includes two criteria: “In the first place, the bystander saves his five by making something that threatens them instead threaten one. Second, the bystander does not do that by means which themselves constitute an infringement of any right of the one’s.” However, in her later paper (Thomson 2008), she changes her mind with reference to a further variation of the scenario with a total of three options for the bystander including the choice between the death of five, the death of one, or self-sacrifice. She now agrees with Foot’s initial proposal and holds that it is not permissible for the bystander to flip the switch.

13 Of course, it is not simply up to the programmer which way to go; manufacturers or states and the corresponding laws will dictate what is permissible to program, and in this respect the programmer is not the decision maker at all (cf. Nyholm/Smids 2016, 1281 f.). For the sake of argument, I will leave these questions aside.
role in trolleyology. What matters instead is the following: If there is a morally significant difference between driver and bystander, the programmer has to be paralleled with the bystander. Both bystander and programmer did not start the vehicle; they are in a certain way outside the situation. Neither of them has initiated a process leading to the death of five in case of non-intervention, neither of them maneuvered the vehicle into a dangerous situation. Of course, there would be no autonomous vehicle without a programmer. But the case of the bystander would not be judged differently even if he had been involved as an engineer in the construction of the trolley. It does not make any difference whether someone was part of the construction of the vehicle or not. Similarly, the fact that the programmer has not only been involved in the construction of the trolley, but also in the decision-making process of where to steer the vehicle, does not change the game; for the bystander also faces a decision to make. Hence, it is also negligible at what point in time the decision is made (obviously, the programmer makes his decisions way ahead of a possible accident). Consequently, the role of the programmer is that of the bystander. Nota bene: This does not imply that the difference between driver and bystander is morally relevant at all. Assuming for now that this difference is relevant, though, we are now in a position to address the main question of this paper: Would Kant agree to let the bystander (and hence the programmer) flip the switch?

3 Kant’s Solution to the Trolley Problem

Let me begin with a comment on possible conflicts of duties. According to Foot, the trolley problem is a situation in which we find ourselves faced with such a conflict. However, for Kant, such conflicts of duties are not possible. This much is clear, I think, but it is not evident why this is so. It seems as if Kant argued for the impossibility of such conflicts by virtue of the very essence of what duties are: Duties obligate with necessity, and therefore, two duties cannot contradict each other. This seems to be a rather formalistic approach to the problem. Yet it

---

14 If at all, psychological pressure can only be relevant in legal contexts when it comes to the question of whether someone who broke the law in a case of (possible) exculpatory necessity (Notstand) should be punished; cf. Wörner (2019, 45 f.).

15 For the impossibility of conflicting duties, cf. Kant’s Doctrine of Right, MS: 224. – Since, strictly speaking, a dilemma involves two conflicting binding rules in a specific situation and conflicts of duties are not possible, the trolley situation, according to Kant, cannot be understood as a dilemmatic situation in a strict sense. Even though duties cannot contradict, according to Kant, “grounds of obligation” can. Unfortunately, Kant says next to nothing about what grounds of obligation are
is possible to develop a further argument on Kantian grounds. To see how, let us recall the different kinds of duties he discusses: Roughly speaking, there are wide duties, i.e., duties which come with a certain kind of latitude in obeying them, and narrow duties without latitude.\textsuperscript{16} Since wide duties only command maxims\textsuperscript{17} and no particular acts, a conflict between wide duties is not possible. For instance, the wide duty of benevolence cannot contradict the wide duty of gratitude in a situation in which one can either carry out an act of benevolence to one person or an act of gratitude to another; both wide duties of benevolence and of gratitude have latitude that enables one to carry out the specific act of benevolence (or gratitude) at another time and another place. Furthermore, since narrow duties are negative duties only commanding omissions, there can be no conflict of narrow duties either; one can both and at the same time obey the duty not to murder someone as well as the duty not to steal.\textsuperscript{18} And since, lastly and generally speaking, both reasons hold (wide duties have latitude and narrow duties only tell us not to do something), there can be no conflict between a wide and a narrow duty either.\textsuperscript{19} Now one could think that the trolley problem strongly supposed to be and in which sense conflicts between them can arise. There is no consensus in the literature about these questions either; possibilities range from bare facts to \textit{prima facie} duties and a combination of both. According to Timmermann (2013, 48), “a ground of obligation arises when an agent correctly applies an ethical principle to a concrete case.” “This ground,” he goes on, “though genuine, can still fail to produce an actual obligation if the agent lacks the means to further the ethical end in question, e.g. because of some physical, psychological, or moral impossibility”. Timmermann (2013, 43–50) holds that conflicting \textit{grounds of obligation} are only possible within the realm of ethical duties (for a different position, cf. McCarty 1991, 69). However, he does not discuss the possibility of ethical duties being duties of omission, especially not duties of respect (I will get back to this later). For further analysis, cf. McCarty (1991) and Adkins (1999), and, in a more general approach, Hill (2002). – Note that the trolley situation is also not a case of \textit{moral uncertainty}: The alleged conflict can be resolved, or so I shall argue. I shall present a conclusive Kantian answer to the question of what to do. In doing so, I will not switch from one level of argumentation (what is morally obligatory in a specific situation?) to another (what ought we to do when the answer to the first question is uncertain?); for an overview regarding such a strategy of distinguishing different levels, cf. Bykvist (2017).


\textsuperscript{17} The term \textit{maxim} is of course a highly problematic one. For this purpose, I take a \textit{maxim} to be a non-normative, expressive principle of an agent’s practical self-determination.

\textsuperscript{18} Cf. Timmermann (2013, 45). According to Joerden (1997, 44) and from a legal point of view, it is not undisputed that there can be no conflicts between negative duties. According to McCarty (1991, 70 f.), there can be “conflicts of grounds of obligation involving only perfect duties” when it comes to the duty of civil obedience or the duty to keep promises on the one hand and the duty not to lie, for instance, on the other hand.

suggests that Kant must be wrong, after all, in claiming this impossibility of ethical conflicts – for the trolley problem shows, it might be argued, that there indeed can be conflicts between duties. But as we shall see, this conclusion would be drawn too quickly.

How would Kant analyze the trolley situation in his own terms, provided that no conflicts of duties are possible? Given this assumption, we cannot think of the bystander as having to choose between obeying two different narrow duties, that is, between the narrow duty not to kill five and the narrow duty not to kill one. Otherwise, the bystander would not be able to obey both duties and so there would be a conflict of duties. Rather, it is possible to interpret the trolley situation, with Kant, in two ways:

(i) The bystander has to choose between obeying two different wide duties: saving the five and saving the one. Given the latitude of wide duties, both actions are permissible; given their permissibility, no conflict is involved (I call this the wide/wide interpretation).

(ii) The bystander has to choose between obeying a wide duty and a narrow duty: saving (the five) and not killing (the one). Given the latitude of the wide duty (to save others) and the narrow and negative character of the other duty (not to kill), it is morally obligatory to obey the latter, i.e., not to kill the one. Given the latitude of the wide duty, no conflict of duties arises. For even though it is, generally speaking, a duty to help others, there is no obligation to save the five in this particular situation, given the narrow duty not to kill the one (wide/narrow interpretation).

With regard to the possibly conflicting “grounds of obligation” (MS: 224) already mentioned, one might wonder whether one could indeed interpret the trolley situation as a choice between not killing the five and not killing the one. Although there is no conflict of duties (for the general duty not to kill as such does not conflict with any other duty), there could be different grounds of obligation concerning this duty not to kill, i.e., the five, as one such ground, and the one as the other. I will not pursue this line of argumentation. Firstly, because it is not clear at all what those grounds of obligation are. This interpretative problem, however, is no reason to worry. In fact, secondly and for reasons that will become clear below, it is not sensible to claim that the bystander kills the five by doing nothing (see below, 200–202). Thirdly, within Kantian ethics, it is never possible to interpret a choice as one between killing and killing. Since to kill innocent persons is strictly forbidden, one has to accept one’s own death before killing someone else (I will get back to this point in detail; cf. also the gallows man example in KpV: 30).

Formally speaking, there is one further possible interpretation of the situation, namely the narrow/wide interpretation (rather than wide/narrow): The bystander would then have to choose between not killing the five (narrow) and saving the one (wide). Since this interpretation is highly implausible, I will leave it aside.
Option (i), the *wide/wide interpretation*, is not very promising. The bystander’s two possible actions would be described as follows. Either the bystander *does* flip the switch in order to obey the wide duty to save the five or the bystander *does not* flip the switch in order to obey the wide duty to save the one. Note that it is quite correct to describe the bystander’s action to pull the switch in order to *save the five* as an action obeying a wide duty. However, to say that the bystander *saves* the one on the track by not flipping the switch is implausible, for the one, at least in a strict sense, is not in danger at all before the bystander enters the scene.\(^{22}\) It is not that the one on the track somehow is in danger of losing his life, and *then* the bystander shows up and has the option of doing away with that danger. If there is a danger at all, it arises *because* of the bystander’s possible thought of actually pulling the switch; but there is, as it were, no danger *ante cogitationem*.\(^{23}\) Also note that it would be of no help (even if it might be correct) to say that there are not *two* wide duties in this situation but only *one*, i.e. the general wide duty to help others. For if this is correct, this one wide duty allows the bystander to choose freely between flipping the switch or not. If he pulls the switch he chooses to help the five; if he does not pull the switch, he chooses to help the one. However, the problem remains even under this interpretation: It makes no sense to say that the bystander *helps* the one by not flipping the switch because the one person is not in need of help, to begin with. To conclude, it is not the case that the bystander has to choose between helping one and helping five.\(^{24}\)

\(^{22}\) According to Kleingeld (2020, 219 f.), the agent has a choice between *saving* and *saving*. However, her line of argument goes into another direction (cf. below, 205–206).

\(^{23}\) For the reasons just given, it makes little sense to argue that all six people, the five and the one, together make up *one* endangered group. (Also note that German law rules out the possibility of singling individuals out to die for the sake of the others even in cases of a “Gefahrungsgemeinschaft” – in which without intervention all will die, cf. Wörner 2019, 48). Of course, *in some sense*, everybody on the tracks is in *some* danger even when no trolley is near. But this general kind of danger is different from the concrete danger the five are in or the concrete danger the one is in *after the bystander comes around*.

\(^{24}\) Joerden (1997, 45ff.) describes the following case: A father has to choose between saving his drowning son A and saving his drowning son B; he cannot save both, and if he does nothing, both will die. In this case, Joerden argues, the father is only faced with one duty indeed, namely the wide duty to help his children (or, generally speaking, to help others). Since he cannot save both, to save both cannot be a duty. According to Joerden (1997, 47), the father is obliged by one duty, but two “grounds of obligation”. – For the reason mentioned above, we cannot interpret the bystander’s situation as similar to this kind of situation. Rather, cases similar to that described by Joerden are so-called *Taurek cases* (in which only duties to help others are relevant); I will get back to these later.
Thus, from a Kantian point of view, option (ii), the wide/narrow interpretation, is the most sensible reconstruction of the bystander’s situation. Since the latitude of the wide duty (to help others) allows one to refrain from carrying out the wide duty in this particular case, the narrow duty overrides the wide duty. Therefore, there is no conflict of duties and it is morally obligatory not to flip the switch and therefore not to kill the one. Put another way: According to option (ii), it is not permissible for the bystander to pull the switch in order to save the five. And of course, this answer to the trolley problem indeed looks very Kantian. Even if we might not like this result from a systematic, non-Kantian ethical point of view, it is the result Kantian ethics leads to.

This is the basic outline of the Kantian solution to the preceding trolley problem. I will now reject three objections and then adduce three additional arguments supporting this interpretation. In the course of this, the Kantian solution will further be unfolded.

4 The Wide/Narrow Interpretation: Three Objections

All three objections aim to show that the wide/narrow interpretation of the bystander’s situation is wrong. The first two do so by denying the wideness of the duty concerning the five, although with different arguments. The third objection argues that the duty concerning the one is not a narrow duty. I will rebut all three objections.

Objection 1) It is not a wide duty to save the five because there is no latitude. Since the bystander has no latitude in helping the five, as posited by the objection, there is no wide duty to help but only a narrow one. The bystander cannot choose between helping the five now or tomorrow, to a greater or smaller extent, in this way or another. Given the typical trolley situation, there simply is no latitude in obeying the (allegedly wide) duty to help the five; if the bystander does not flip the switch, the five will die. Since wide duties require latitude and here there is no latitude, there can be no wide duty to help in this case, and therefore the wide/narrow interpretation cannot be correct. – This objection overlooks the difference between the general wide duty to help others and the alleged specific obligation to save the five on the track. Even though the bystander has the wide duty to help others in general, he is not obliged to
save the five in the trolley situation.\textsuperscript{25} Hence, there is latitude in obeying the general wide duty to help others – it is the latitude of the wide duty that indeed does make it morally permissible and even necessary to refrain from helping the five. Since a wide duty does not command specific actions but only maxims, it is not obligatory for the bystander to help the five. Once again: The point is not that there is no wide duty; there is, as there always is the wide duty to help others in need, but this duty does not apply given the specific situation. Thus, the bystander’s situation includes both a wide as well as a narrow duty; they do not conflict, but the narrow duty trumps the wide duty.\textsuperscript{26} (Furthermore, saving someone’s life certainly cannot be a narrow duty in principle, for then we would be required to constantly save whosoever’s life.)

However, matters do get more complicated when we dig a little deeper into Kant’s distinction of duties. There are at least two problems: First, although Kant claims that all ethical duties (which the \textit{Doctrine of Virtue} allegedly consists of) are wide duties that come with latitude,\textsuperscript{27} we do find negative duties (perfect duties) in the \textit{Doctrine of Virtue} as well, such as the prohibition of suicide or lying, for instance, and as negative duties, they seem to lack any latitude. Given the rather strong and repeated claim of Kant’s that ethical duties are wide duties, there has to be a way to think of those negative duties in the \textit{Doctrine of Virtue} as wide duties, too – or else Kant’s distinction between wide and narrow duties on the one hand and between the duties of the \textit{Doctrine of Right} and the \textit{Doctrine of Virtue} on the other would make no sense whatsoever. Therefore, we must argue that not all negative duties are narrow duties.\textsuperscript{28} Very briefly: There is (i) not only a casuistry for duties of commission (imperfect duties), but also for negative duties\textsuperscript{29} – and since casuistry, if taken seriously, requires latitude, negative duties that allow for a casuistry seem to be wide. (ii) Kant repeatedly stresses that ethical duties as such, \textit{i.e.} including negative or imperfect duties, only command maxims, which also implies that they are wide

\textsuperscript{25} For the distinction between \textit{duty} and \textit{obligation}, cf. also Timmermann (2013, 42 f.) and McCarty (1991, 68). \textit{Duty} refers to the general duty before its application, \textit{obligation} means the duty applied to a specific situation.

\textsuperscript{26} I will sometimes speak of the \textit{wide duty concerning the five} even if there is no duty concerning the \textit{five} in the strict sense but only the general duty to help others.

\textsuperscript{27} Cf. Kant’s \textit{Doctrine of Virtue}, MS: 390.

\textsuperscript{28} It is tempting to understand all legal duties as narrow duties and thereby as negative duties (cf. Mieth/Bambauer 2018, 116). However, with regard to the \textit{Ulpian formulas} (MS: 236 f.), things might be more complicated. The third of these formulas, for instance, is not a negative one: “(If you cannot help associating with others), \textit{enter} into a society with them in which each can keep what is his \textit{(suum cuique tribue)}.”

\textsuperscript{29} Kant speaks of “\textit{casuistry}” (MS: 411) as well as of “\textit{casuistic questions}” (MS: 423). For \textit{casuistic questions} regarding the duty not to commit suicide, cf. MS: 423 f.
duties with latitude. Kant at least at one point puts emphasis on his claim that negative duties can indeed be of a “wide obligation.” – In light of these distinctions, I take the duty not to kill other persons (which applies regarding the one on the track) to be a narrow duty, that is, as a legal duty of omission. Given that all legal duties are also “indirectly ethical“ duties (MS: 221), one can also consider it as a negative, ethical duty (parallel to the duty not to kill oneself).

The second problem with Kant’s system of duties is the following. One might think that even for Kant wide duties must be able to be narrowed down, so to speak. For instance, someone who has the general wide duty to help others has to help a drowning child if nothing else speaks against the act of helping; there is no latitude. The crucial point, however, is that wide duties remain wide even if in a specific situation there is only one action that is to be carried out. The fact that the wide duty, in a specific situation, must be realized through a specific act does not dispute its wideness, because the wide duty remains open to other acts in other situations. In any case, the trolley situation is by no means similar to that of the drowning child. For there is a reason not to save the five (to wit, the one on the track), whereas there is no reason whatsoever not to save the drowning child. To save someone’s life is to fulfill a wide duty, and this is also true in the bystander’s situation. The wide/narrow interpretation remains undefeated.

Objection 2) Not flipping the switch really is killing the five. According to the second objection, the wide/narrow interpretation is wrong since not to pull the switch is tantamount to killing the five, and since killing is strictly prohibited, it cannot be a

30 Cf. Kant’s Doctrine of Virtue, MS: 410, for instance.
31 Cf. Kant’s Doctrine of Virtue, MS: 394.
32 Perhaps, one could say, as Mieth/Bambauer (2018, 122, my own translation) suggest, that a wide duty “articulates both an imperfect and a perfect duty, depending on the specific context: It is a perfect duty only if the life of the other is at stake, but otherwise meritorious in the sense of morally desirable”.
33 With regard to situations in which someone could save a child’s life by committing a minor crime such as stealing something from a supermarket, for instance, Mieth/Bambauer (2018, 117) maintain the following: One can either argue that the wide (ethical) duty to help others turns into a narrow (legal) duty, or that the wide ethical duty trumps the legal duty. I would submit, however, that both options seem to run afoul of Kant’s system of duties. As opposed to supermarket situations, it seems quite obvious that killing the one cannot be considered a minor crime; also, Kant would not even allow minor crimes to be committed in order to bring about something good. The question of whether minor crimes could be exculpatory on Kantian grounds is difficult to answer. Kant holds: “Yet there could be no necessity that would make what is wrong conform the law.” (MS: 236) Unpunishable actions (cf. MS: 236,1 f.) are to be distinguished from actions that might be exculpatory. I cannot get into this here.
wide duty to save the five. – This objection is not convincing either. Not flipping the switch is not tantamount to killing the five, at least not in a morally relevant sense. I shall argue that this is not only true for the bystander but also for the driver of the trolley. Although most philosophers like Foot or Thomson typically assume – and do so without much ado – that the driver *kills* the five by not pulling the switch because he started the trolley (and here *killing* is supposed to be morally relevant killing, i. e., killing which violates negative duties), this assumption is flawed or by no means indisputable. Rather, the fact that the driver started the trolley is irrelevant. Let me draw an analogy: Think of someone who starts his non-autonomous car and drives up to the main road with regular speed. Suddenly, a young boy riding his bike crosses the street out of the blue, leaving the driver no chance whatsoever to swerve; the boy dies from the crash. Did the driver *kill* the boy? Given that the driver wasn’t speeding, that he wasn’t under the influence, given that he drove his car cautiously and so on, we would not say that he *killed* the boy, at least not in a morally relevant way. Of course, the boy is dead (he surely got killed), and of course there is a causal relation between the driver and the boy’s death; the boy would not be dead if the driver had not gotten out of bed that morning. But the driver neither killed the boy intentionally or deliberately (he did not want to see the boy dead) nor acted negligently. So even if on occasion legal scholars or philosophers indeed say that A killed B although it was in no way A’s intention or fault, such a killing is not a *morally relevant* killing. (As a matter of fact, it seems to me that we should not say that the driver killed the boy even if we mean this in a morally neutral way; rather, we should say that the boy died in an accident.) And so just like the driver of the car, the driver of the trolley also does

---

34 See above, footnote 11. Thomson (1985, 1398) states the following: “[. . .] we might well wish to ask ourselves what exactly is the difference between what the driver would be doing if he failed to turn the trolley and what the bystander would be doing if he failed to throw the switch. As I said, the driver would be driving a trolley into the five; but what exactly would his driving the trolley into the five consist in? Why, just sitting there, doing nothing! If the driver does just sit there, doing nothing, then that will have been how come he drove his trolley into the five. [paragraph] I do not mean to make much of that fact that about what the driver’s driving his trolley into the five would consist in, for it seems to me to be right to say that if he does not turn the trolley, he does drive his trolley into them, and does thereby kill them.” Yet it is obvious that this is not really an answer to the question posed by Thomson herself. She does not account for the *difference* between the driver and the bystander; she simply assumes that the driver commits an act of killing whereas the bystander does not.

35 A similar proposal is made, albeit very sketchily, by Mannino/Mukerji (2020, 106).

36 I am grateful to Reinhard Merkel for helpful comments regarding the legal context. – Thomson (1991, 289) holds: “[. . .] if an event that consists in the fall of Y on X kills X, then it follows that Y killed X, whatever Y may be”.

not kill the five in a morally relevant way by doing nothing; he neither wants to kill the five nor is it his fault that something went wrong with the trolley. At first sight, the difference between the driver of the car and the driver of the trolley appears to be that the latter has an alternative, i. e. he could turn the trolley, whereas the former has no alternative (or so we assumed). But it is important to see that the driver of the trolley has, on second thought, no alternative either. For the only thing he could do is something which violates a narrow duty. Within Kantian ethics, to violate a narrow duty is not an option one may consider if one wants to do what is morally right. It is not permissible for the driver to turn the trolley; instead, he is obliged not to turn the trolley in order not to kill the one. Unlike the driver of the car, who is not even in a position to carry out an action that would save the boy, the driver of the trolley is (physically, as it were) in a position to carry out an action that would save the five; but this action is prohibited. It is important to stress that this is exactly what he is doing: he does not turn the trolley in order to fulfill his duty not to kill the one. Since he is not morally free to do otherwise, what he does cannot be described as killing the five (for again, a duty can only be neglected when the person who acts has a permissible alternative to choose). Note, too, that if there were no switch to be pulled in the first place, so that the five would inevitably die, neither the driver nor the bystander would kill the five (in a morally relevant way). Differently from what Foot assumes, the driver of the trolley does not have a choice between killing and killing; by not turning the trolley, he does not kill the five in a morally relevant way. Since this is true concerning the driver, it is all the more true that the bystander does not kill the five by doing nothing. This leads to an important point: There is no morally relevant difference between the driver of the trolley and the bystander. Neither has a choice between killing and killing: rather, both must choose between saving and killing.37 Thus, to come back to the role of the programmer, it does not make any relevant difference whether we parallel the programmer of an autonomous vehicle with the driver of the trolley or the bystander. In Section 2, I pointed out that this question of whether the programmer is to be identified with the driver or the bystander is a question that deserves serious attention, for it could have made a difference. We now see, however, that the difference is actually irrelevant. Both driver and bystander have a choice between saving the five and killing the one. Not saving the five is not tantamount to killing the five. The wide/narrow interpretation remains undefeated.

37 Kamm (2016, 58), for instance, also holds that the question of how the trolley was turned is more important than the question of who turned the trolley.
Objection 3) It is not a narrow duty not to flip the switch because flipping the switch does not mean to kill the one. Given that whoever pulls the switch does not want to see the one dead but save the five (and would, if possible, do the latter without doing the former) – the objection goes – to flip the switch is not tantamount to killing the one (again, in a morally relevant way). So after all, it might as well be permissible to pull the switch because pulling the switch does not violate the narrow duty not to kill (since it does not mean to kill the one). – However, the opposite is true: To flip the switch does indeed mean to kill the one. One alternative to saying that one has a narrow duty not to kill the one would be to say that one has a wide duty to save the one. However, this option is null since the one on the track is, as described above, not in danger in a strict sense before someone thinks about turning the trolley; therefore, saving does not fit the situation. Now, and most importantly, the most obvious way to reinforce the thesis that flipping the switch is not tantamount to killing the one is the Doctrine of Double Effect (DDE). Suppose someone argues the following: Since we are confronted with a situation in which, firstly, the act itself (flipping the switch) is permissible, secondly, the negative effect (the death of the one) is not directly intended, thirdly, the death of the one is not a means to saving the five, and, finally, the negative effect stands in an adequate relation to the positive effect, the DDE applies and therefore it is permissible to flip the switch to save the five even though it will lead to the death of the one.\(^{38}\) – However, there is no room for the DDE in Kant’s ethics. Even if we ignore the severe systematic problems of the DDE,\(^{39}\) it is not possible to deploy the principle in accordance with Kantian theory for two reasons. (i) As far as I can see, there is no direct reference to the DDE in Kant’s (published) writings; if it did play any substantial role in Kant’s ethics, one would expect it to show up at least in his main works. While self-defense is a classical Aquinian example for implementing the DDE, Kant does not refer to the DDE when he discusses that topic (according to Kant, self-defense leading to the aggressor’s death is only permissible when someone is confronted with an unjust aggressor; it is not permissible to kill an innocent person to save one’s own life –

---

\(^{38}\) For this classical version of the DDE, cf. Mangan (1949, 43). – Note that the DDE, strictly speaking, is not meant to show that the killing of a person can be permissible. Rather, if at all, the DDE is supposed to show that it can be permissible to carry out an otherwise morally neutral action (in this case, flipping the switch) leading to the death of a person.

\(^{39}\) Among other things, it seems impossible to determine what exactly the so-called act in itself is (since every act must have an end, according to Kant, this problem might especially arise for Kantian ethics). Furthermore, the relevance of the underlying distinction of intended and merely foreseen harm can be questioned, of course. For the last point, cf. for instance Thomson (1991, 292–296).
I shall come back to this below).\(^{40}\) (ii) The \textit{DDE} is in contrast with the very spirit of Kantian ethics; the fourth condition is especially unsuited to Kant’s thoughts. There never can be an adequate relation between good and bad effects when persons, i. e. human beings endowed with dignity are concerned. One cannot weigh the worth of five persons against the worth of one person (I shall elaborate on this thought below). All in all, pulling the switch does indeed violate the narrow duty not to kill the one. The \textit{wide/narrow interpretation} remains undefeated.

\textit{Excursus: The loop case.} Thomson (1985, 1401–1403) created the \textit{loop case} to show that the trolley problem cannot be solved by referring to the Kantian principle \textit{not to use another person merely as a means}. In the loop case, Thomson says, it is permissible to flip the switch although the one on the other track, the large man who gets killed as a consequence of pulling the switch, is used as a means to save the five (because the track with the one and the track with the five circle back in a loop so that the five would be killed regardless of the switch’s position, were it not for the large man whose body could stop the trolley). – Let me take up three points here: First, it is important to see that in debates initiated by the loop case, two elements are somehow intertwined: the \textit{DDE} and the Kantian principle \textit{not to use another person merely as means}. As we have seen above, the Kantian principle is, at least in some sense, one element of the \textit{DDE}, to wit, that the bad consequence (the death of the one) must not be a means to save the five. Given this conjunction of the \textit{DDE} and the Kantian principle, the following is true: To hold that it is \textit{permissible} to flip the switch in the loop case (even though the death of the one is a means to save the five) implies that the Kantian principle \textit{not to use another person merely as means} is violated or rather refuted – and \textit{a fortiori} so is the \textit{DDE}.\(^{41}\) Yet the fact that the Kantian principle is indeed (in some sense) part of the \textit{DDE} does \textit{not} show that Kant embraces the \textit{DDE}.

\(^{40}\) There is no consensus concerning the question of whether the \textit{DDE} does go back to Thomas Aquinas (cf. again Mangan 1949). For the Aquinian example of self-defense, cf. his \textit{Summa Theologica} (II–II, Qu. 64, Art. 7). Cf. Kant, MS: 235 f., TP: 300. – According to Byrd, Kant refers at one point (MS: 336) to the so-called “\textit{actio invita (the reluctant act)}” (Byrd, 2015, 2299, my own translation) which, he says, is discussed by Achenwall and goes back to Aristotle’s thought on mixed acts. However, since Byrd only refers to this single passage of Kant’s, this linkage seems to require some more evidence. Also, the context of this utterance of Kant’s is the topic of duels, that is, a context in which the person to be killed is not innocent but has harmed the other. The agent’s innocence in moral relations is crucial here; I will get back to this shortly.

\(^{41}\) Thomson does not refer to the \textit{DDE} in her discussion of the loop case. However, she does reject the \textit{DDE} (see Thomson 1991, 292–296). – Kamm (2007, 91–129) famously tried to save the \textit{DDE} at least in some sense with the introduction of the \textit{Principle of Triple Effect} and her distinction between \textit{doing something in order to} and \textit{because of} (which others were quick to
For, as mentioned above, the DDE contains other elements which Kant indeed rejects.\footnote{Here I concur with Parfit (2017, 382 f.), who makes exactly the same point; he also shares the opinion that the DDE is of no relevance for Kant. In discussions about the DDE, Kant is rarely mentioned and only with regard to the principle not to use another person merely as means. For instance, in his recent book on the DDE, Černý (2020, 99) mentions Kant once, but only to refer to the prohibition of using someone merely as a means.}

Second, Pauline Kleingeld recently argued that, as a matter of fact, turning the trolley in the loop case does not necessarily go along with using the one merely as a means even on Kantian normative grounds. It all depends, she argues, on how the bystander relates to the one in his thinking about how to act. Unlike the case with the large man and the bridge, in which pulling the switch clearly implies using the one merely as a means, in the loop case – she says – it is only possible but not necessary to use the one merely as a means. Rather, the bystander could set himself the end of saving the five, but not by using the one as a means, although he knows that the one would ultimately stop the trolley. From this perspective, the one does not “enter into her [the bystander’s] reasoning as a means” (Kleingeld 2020, 220). However, Kleingeld’s argument seems flawed. For one thing, she describes the decision at stake as one between saving and saving and not as one between killing and killing (as Foot and Thomson do) or saving and killing (as Kant does, or so I claim).\footnote{Kleingeld (2020, 219 f.).} For another, and most importantly, it is quite hard to see what her argument actually is. She admits that the bystander knows that it is the one who will stop the trolley. But simply saying that this functional role of the one does not enter into the bystander’s reasoning does not prove that this way of reasoning is sensible. If the agent knows the one is, objectively speaking, the means to save the five (if he were not on the track, switching the track would not make any sense), just saying that nonetheless and somewhat mysteriously, the one does not enter into the bystander’s reasoning as a means is neither a good account of what actually happens nor a good representation of what goes on in the bystander’s mind. Kleingeld describes what she thinks is the morally permissible maxim of the bystander in the loop case as follows: “[. . .] either I let the trolley continue towards the five, which will save one life, or I divert it towards the heavy man, which will kill him but save five lives” (2020, 220). Since there is no mention of the large man as a means, he is not used as a means from the
agent’s perspective, she says. However, it is not clear why we should not be able to describe the bystander’s maxim in the case with the large man and the bridge in a similar way. For we could very well describe the bystander’s maxim in the footbridge case as follows: *either I let the trolley continue towards the five, which will save one life, or I throw the large man down the bridge, which will kill him but save five lives.* Here, too, there is no mention of the large man *as a means*, but from this it does not follow that it makes sense to say that from an internal perspective, he is not used as such. Put another way: It is not, at least not generally speaking, up to an agent’s reasoning whether he does use someone (merely) as a means or not. Without there being a man on the loop’s tracks, pulling the switch would not make any sense.

Third, even though loop discussions are, of course, quite enlightening, they are irrelevant for the question to be answered in this paper. For reasons that have been already laid out and will be further developed in due course, Kant would not permit flipping the switch, neither in the original trolley case (both for the driver and for the bystander) nor in the loop case. Indeed, it looks as if, for Kant, there is a short-cut answer to what I have called the prevalent trolley problem (the question of why turning the switch is permissible in some cases but not in others): this problem does not arise to begin with, given Kant’s answer to what I have called the preceding trolley problem (the question of whether it is permissible to turn the trolley in the original case). One may not and never pull the switch; therefore, the question of why one may pull the switch in some cases but not in others simply does not arise.

### 5 Unfolding Kant’s Solution: Three Additional Arguments

Having cleared these three objections out of the way, it seems not only possible but quite plausible to interpret the trolley situation in terms of the *wide/narrow interpretation*, which implies that whoever is in charge of the switch is obliged not to turn the trolley. I will now present three additional arguments to support this view. This will also help to further unfold the Kantian solution to the preceding trolley problem.

*Additional argument 1) The maxim not to turn the trolley is universalizable on the basis of the principle of universalization.* The categorical imperative (CI) is the central tool of Kantian ethics and supposed to be a procedure for testing the
universalizability of maxims (if a certain maxim is universalizable, the action in question is at least morally permissible). I shall now show that the maxim not to pull the switch is universalizable. Since testing a maxim presupposes having phrased an appropriate maxim in the first place, we have to ask: What maxim adequately describes the act of not flipping the switch? Of course, this is a question that touches on a lot of subjects and further questions, so I have to focus on some well-chosen aspects.

Let me begin with two questions, one rather basic and one somewhat specific: What is Kant’s account of acts? And since the act in question is not to pull the switch: Are omissions acts? Very briefly, the answer to both questions is this: According to Kant, a “deed” (MS: 223) is a free and accountable act under the moral law which as such can only be performed by free, autonomous subjects capable of active and passive moral obligation (persons). A person acting in this sense is the author of a deed’s results and therefore accountable for them. Since the CI is supposed to be (also) a procedure showing which duties we have to fulfill, and since a duty can be positive (a duty of commission, i. e. a duty to do something) as well as negative (a duty of omission, i. e. a duty not to do something), omissions can also be deeds, at least when it comes to negative duties.

So to refrain from doing something that is prohibited by the moral law is a deed: “All moral omissions are negative actions and therefore not a lack of actions: but real actions realiter opposed to the positive actions” (PPH: 9). This understanding of omissions fits quite well with Kant’s concept of what it means to obey the moral law. To act from duty means to act as our noumenal self tells us to and thus to resist our selfish inclinations. Not to perform an act we feel inclined to carry out means doing something, to wit, actively resisting these inclinations out of respect for the moral law. If someone feels the strong need to lie to his

---

44 Cf. in more detail, Schmidt/Schönecker (2017, 136–148).
45 “An action is called a deed insofar as it comes under obligatory laws and hence insofar as the subject, in doing it, is considered in terms of the freedom of his choice. By such an action the agent is regarded as the author of its effect, and this, together with the action itself, can be imputed to him, if one is previously acquainted with the law by virtue of which an obligation rests on these” (MS: 223).
46 As stated above, Kant distinguishes between wide and narrow duties. Wide duties can be either positive duties (as the duty to be benevolent; cf. §§ 29–31 of the Doctrine of Virtue) or negative duties (as the duty not to kill oneself, for instance; cf. § 6). We have only duties to persons, either to oneself or to other persons (Kant also speaks of duties with regard to nature, animals and God which are really duties to oneself; cf. §§ 16–18 of the Doctrine of Virtue).
47 Cf.: “By contrast, imagine a human being who fails to help someone whom he sees in distress and whom he could easily help. There is a positive law to be found in the heart of every human being, and it is a law which is present in this man’s heart as well; it commands that we
spouse but decides not to lie because it would be the wrong thing to do, he acts out of respect for the moral law. Thus, omissions are or at least can be deeds, and not to flip the switch clearly is an omission, and as such a deed.

To formulate an adequate maxim we must clarify what exactly the deed under consideration, understood as an omission, consists in: Is it (a) simply not pulling the switch without further qualification? Is it (b) not flipping the switch and thus letting five die (we already ruled out that whoever does or does not turn the trolley kills the five in a morally relevant way by not pulling the switch)? Or is it (c) not flipping the switch and thus not killing the one (recall that pulling the switch is tantamount to killing the one)? It is undisputed that deeds cannot only consist in the physically (not) carried out movement, and since every deed, according to Kant (MS: 385), has an end which essentially constitutes the deed by the maxim the subject is determined through, it cannot be option (a). Thus we are left with options (b) and (c), and then the question is: What is the end of a person who does not turn the switch? Of course, that depends precisely on which end she sets. Given the typical trolley situation, it is taken for granted that it is nobody’s end to let the five die (they are nobody’s enemies, or whatever). Therefore, option (b) is eliminated as well and option (c) is the correct specification of the maxim: “I do not flip the switch in order not to kill the one”.48 Again, as an act of omission, the end is to prevent something from happening (killing the one); killing the five is not part of the deed. Of course, the maxim just formulated looks quite unproblematic. But what about the maxim “I do not pull the switch in order not to kill the one knowing that five will die”? Is this maxim also universalizable? Or, to put it another way: Even if it is, generally speaking, true that not to kill someone is a narrow duty, could it be permissible or even morally obligatory to kill someone in a very specific situation in which the act of not killing elsewise would lead to even more harm (the death of five)?

love our neighbour. In the present example, the law must be outweighed. For this omission to be possible, it is necessary that there should be actual inner action arising from motives.’ This zero is the consequence of a real opposition. And it really does initially cost some people a noticeable effort to omit performing some good, to the performance of which they detect within themselves positive impulses. Habit facilitates everything, and this action is in the end scarcely noticed any longer” (NG: 183). And: “Both omission and commission are actus of freedom. In the moral sense, they are both called acts, although omission is not an act in the psychological sense. In the moral sense, everything that has a motive [Bewegungsgrund] is an act. Omitting something for a moral motive [Bewegungsgrund] is just as much an action as doing something; for example, not taking revenge on someone who has offended us is really more difficult than taking revenge” (PPH: 128).

48 From another perspective, we already came to the same result above; see p. 202.
This leads to a notorious allegation that has often been brought up against Kant’s ethics – his deontological rules were too rigorous. Let us have a look at the prohibition of lying and a well-known example related to it. Suppose Peter, who lives in Germany in 1942, harbors ten Jews and lets them hide in his attic. One night, some SS-men ring the bell, searching and asking about Jews. Is it permissible for Peter to lie to the SS-men, telling them he does not know any Jews at all? Since for Kant lying is prohibited in each and every case, the answer seems to be no. But this, the critics say, cannot be right; it cannot be right to let the Jews die if the only thing to do to avoid it is to lie to evil Nazis. But regardless of the question of whether it is indeed morally obligatory to lie or not, it is true that for Kant the answer is indeed that we must never lie, without exception. This holds both from a legal as well as from an ethical point of view. As stated in the Doctrine of Virtue, not to lie is an ethical duty to oneself since lying (to others or to oneself) involves “an end that is directly opposed to the natural purposiveness of the speaker’s capacity to communicate his thoughts, and is thus a renunciation by the speaker of his personality” (MS: 429); as famously stated in On a Supposed Right to Lie from Philanthropy, not to lie is “a duty of right” (VRML: 427 fn.) since lying “makes the source of right unusable” (VRML: 426). It is crucial for both the moral and the legal context that the wrongness of lying is independent of the harm to others that might be caused by a lie. With regard to another example, Kant concedes in On a Supposed Right to Lie from Philanthropy:

[. . .] if you have by a lie prevented someone just now bent on murder from committing the deed, then you are legally accountable for all the consequences that might arise from it. But if you have kept strictly to the truth, then public justice can hold nothing against you, whatever the unforeseen consequences might be. It is still possible that, after you have honestly answered “yes” to the murderer’s question as to whether his enemy is at home, the latter has nevertheless gone out unnoticed, so that he would not meet the murderer and the deed would not be done; but if you had lied and said that he is not at home, and he has actually gone out (though you are not aware of it), so that the murderer encounters him while going away and perpetrates his deed on him, then you can by right be prosecuted as the author of his death. For if you had told the truth to the best of your knowledge, then neighbors might have come and apprehended the murderer while he was searching the house for his enemy and the deed would have been prevented. Thus one who tells a lie, however well disposed he may be, must be responsible for its consequences even before a civil court and must pay the penalty for them, however unforeseen they may have been [. . .].

(VRML: 427)

49 However, given that ethical duties are supposed to have latitude (cf. MS: 390), that the prohibition of lying is an ethical duty, and that there is a casuistry concerning lying, things could be more complicated.
Someone who tells the truth is not responsible for bad consequences that possibly might occur; for it might be the case that these consequences change or vanish. On the contrary, someone who does not tell the truth is responsible for all consequences. Again, things could change and his lie could also cause harm:

“It was merely an accident (casus) that the truthfulness of the statement harmed the resident of the house, not a free deed (in the juridical sense)” (VRML: 428).

Lying is wrong – regardless of the consequences.

Can the same be said about killing? Unfortunately, neither the Doctrine of Right nor the Doctrine of Virtue explicitly contain a specific duty not to kill other persons (there is only the prohibition of suicide in § 6 of the Doctrine of Virtue). Here we do not have to address the question of why Kant does not explicitly treat homicide as such. In any event, the moral demerit of killing other persons is not as far-reaching as the demerit of lying. For it is beyond doubt that, on Kantian grounds, murderers have to be punished and even be sentenced to death. Although it is always prohibited to kill innocent persons, this is not true for persons who have committed a severe crime. Since all legal duties are ethical duties, too, to kill innocent persons is always wrong, both legally and ethically. As a matter of fact, Kant explicitly states that it is wrong to kill another innocent person even when one’s own life is at stake: Discussing the famous Plank of Carneades, Kant holds that it is not permissible to push another innocent person off the swimming plank in a situation in which the only way to survive is to stick to the plank. While such behavior is “unpunishable” (unstrafbar), it is “not [. . .] inculpable” (nicht unsträflich; MS: 236) – it is wrong to push an innocent person down the plank thereby causing his death.

50 To some extent, this line of reasoning runs contrary to the trolley cases because in these it is assumed that the five (or the one) will undoubtedly die.

51 This is also what Kant says in the casuistry of the Doctrine of Virtue regarding a slightly different, but similar case: “For example, a householder has ordered his servant to say ‘not at home’ if a certain human being asks for him. The servant does this and, as a result, the master slips away and commits a serious crime, which would otherwise have been prevented by the guard sent to arrest him. Who (in accordance with ethical principles) is guilty in this case? Surely the servant, too, who violated a duty to himself by his lie, the results of which his own conscience imputes to him.” (MS: 431).

52 Cf. Kant’s Doctrine of Right, MS: 332–337.

53 Kant also discusses at least two further cases of killing in this regard. Both concern “the feeling of honor” (MS: 336), firstly in the case of a mother killing her illegitimate child and secondly in the case of duels. These cases are difficult to understand. Nevertheless, Kant seems to say that these acts of killing are wrong and must be punished under ideal legal conditions. The only question that arises is whether the state has reached a level of justice that is close enough to these conditions so that capital punishment is justified.

(note that in a case of self-defense, the situation is different since the other person is not innocent). This is why Kant writes: “[. . .] not to take the life of another who is committing no offense against me and does not even lead me into the danger of losing my life is an unconditional duty” (TP: 300, fn.).

Now let us go back to the trolley case. Since it is presupposed that there are no relevant differences between the persons involved in the trolley case, it is crucial that the one on the track is an innocent person. Also, the life of the person flipping the switch is not at stake, so it is no case of “right of necessity” (MS: 235). This leads to the inevitable conclusion that pulling the switch and thus killing the one cannot be permissible. To kill the one person on the track is wrong, regardless of the consequences.

So far, so good. But we still have to test the maxim not to flip the switch in a strict sense. As we know, there are different formulas of the CI. I will concentrate here on the principle of universalization (“act only in accordance with that maxim through which you can at the same time will that it become a universal law”, GMS: 421). The universalization of the maxim under consideration leads to the following formulation:

It is a universal law that all rational beings in a trolley situation decide not to flip the switch in order not to kill one innocent person, despite knowing that five other innocent persons will die.

As Kant holds in GMS: 424, we have to check for so-called contradictions in conception or contradictions in willing to see whether a maxim is universalizable or not, whereby those different contradictions are said to somehow correspond with perfect and imperfect duties. Can I will without contradiction that it is a universal law that all rational beings must decide not to kill one innocent person even though doing so would save five others? Well, as opposed to the case concerning potential suicide out of selflove and also different from the cases concerning false promises or lying in general, there seems to be no obvious contradiction here. Of course, this answer is not as satisfying as it could be,
but that is due to the fact that the whole procedure of testing maxims is problematic and often said to be malfunctioning.58

Let us think for a moment about the universalizability of the alternative maxim “I do flip the switch in order to save the five, knowing that I will thereby kill one innocent person”. There appears to be a contradiction between killing and the general prohibition of killing innocent persons as described above (which would be circular), or between killing and the required respect for beings with dignity, but I submit that this is not fully convincing either. As we worked out above, to kill innocent persons is forbidden, according to Kant; but that is one thing to say. Another thing to say would be to point to a specific contradiction, where this contradiction is supposed to be the reason for the prohibition in the first place. This too shows that Kant’s idea of testing maxims is burdened with difficulties. However, there is no obvious problem with the universalizability of the maxim not to flip the switch.

Additional argument 2) The maxim not to turn the trolley is universalizable on the basis of the principle of humanity. Kant’s principle of humanity (“So act that you use humanity, whether in your own person or in the person of any other, always at the same time as an end, never merely as a means”; GMS: 429) commands not to use a person merely as a means. At first sight, it is quite obvious that the death of the five is in no way a means not to kill the one and a fortiori the five are not used merely as a means (by the same token, the death of the one is not a means to save the five since the five would also be saved by flipping the switch without there being anyone on the other track). However, things get more complicated once two different interpretations or rather aspects of the formula of humanity are distinguished. First, it is not permissible to use someone only as a means in a direct way (it is not permissible to throw the large man down on the tracks to stop the trolley). In this sense, no one is used as a mere means in the original trolley case, regardless of whether the switch is flipped or not.59 Second, it is not only not permissible to use someone as a means, but it is also not permissible to carry out actions that do not do justice to a person’s unconditional worth, that is, to his or her dignity. In this (second) sense, someone is also used as a mere means when he is not treated with the appropriate

59 However, according to Welzel (1951, 51 f.), the bystander has to pull the switch although he thereby does something wrong (“[. . .] the perpetrator must inescapably accept lesser moral guilt in order to escape greater guilt”). Welzel refers to Kant’s formula of humanity to show that the bystander uses the people on the track as means when he pulls the switch (but still, he claims, it is the right thing to do).
respect he deserves. In the *Groundwork*, after having presented the formula of humanity, Kant discusses four examples, the last two of which undergird this interpretation: It is neither permissible not to cultivate one’s own talents nor is it permissible to refrain from promoting other people’s happiness; in the context of these examples, it is striking that Kant does not speak of not using anyone merely as a means. Rather, he states: “[…] with respect to contingent (meritorious) duty to oneself, it is not enough that the action does not conflict with humanity in our person as an end in itself; it must also harmonize with it” (GMS: 430). Recall that saving or helping others is only a wide duty with certain restrictions: It is not permissible to promote another person’s happiness when the other person’s ends are immoral or when the promotion of the other person’s happiness is, or involves, a violation of a narrow duty. Flipping the switch, however, means violating the duty to respect the dignity of the one, for not to kill is a narrow duty. Therefore, not pulling the switch does not violate the formula of humanity interpreted in that second way either. The respect one owes to the dignity of the five is not neglected if one decides not to flip the switch. (Furthermore, one violates the formula of humanity by attributing a greater worth to the five than to the one; I will discuss this argument below.) The maxim not to pull the switch remains morally permissible.

*Additional argument 3) Numbers are irrelevant: dignity and deontology.* Let us now imagine the trolley situation with different numbers. Suppose there is a choice between watching the trolley hit just one (rather than five) by driving straight on and flipping the switch, leading to the death of one other person. Would Kant concur with pulling the switch in this case? It seems clear to me that he would not, and the reason for this is Kant’s concept of dignity as an absolute value and his deontological approach to ethics. In light of this deontology, it is not possible to offset beings that possess dignity against other such beings. As

---

60 Cf. for this distinction Mieth (2014, 17) and, with reference to Thomas Hill (1980), Mieth (2018, 118–122).

61 Cf. MS: 388 where Kant says that one must only promote the happiness of other beings by making their ends one’s own, provided that their ends are “permitted”.

62 Incidentally, this is also what the German jurisdiction says with a strong reference to Kant, cf. Wörner (2019, 43 f.; my own translation) with regard to a slightly different case: “The justification [of pulling the switch] fails, however, because the good to be preserved [Erhaltungsgut] (life of the persons on the train) does not substantially outweigh the good to be intruded [Eingriffs gut] (life of the three workers). According to the prevailing opinion, human lives cannot be weighed against each other, neither according to their quality (track workers against train passengers) nor according to their quantity (50 against 3); rather, each individual human life stands on its own.”
famously stated in GMS: 428, beings that possess dignity have an absolute value and no prize.\textsuperscript{63} One can weigh them neither against other things, nor against other beings with value; there is no “scale” of dignity.\textsuperscript{64} One cannot sacrifice one person for another, and this is also why one cannot sacrifice one person for five others – \textit{numbers do not count}. Otherwise, the distinction between deontology and consequentialism would blur at least to some extent.\textsuperscript{65}

The impossibility of weighing a certain number of beings with dignity against another number is further illuminated by taking into account so-called \textit{Taurek cases}.\textsuperscript{66} These are cases in which a subject has to choose between saving a relatively small number of people (or even just one person) or saving a larger number of people; it is presupposed that both options only differ by their numbers. Importantly, such cases are different from trolley cases because the subject has to choose between \textit{helping} and \textit{helping}; as opposed to the one in the trolley case, all persons in Taurek cases are already endangered independently of the subject that has to make a decision. For instance, imagine you are at high seas with your yacht and you are informed about five people in distress fifty miles away from you to the north and ten people in distress fifty miles away to the south; you cannot save both the five and the ten. Are you obliged to save the ten \textit{because they are more than five}? John Taurek argues – though against the background of a subjectivist metaethical system very different from Kant’s – that you are \textit{not} obliged to save the ten. Thus, he says, it is permissible, though of course not obligatory, to save the five rather than the ten; as a matter of fact, you should flip a coin to choose which way to go so that everyone involved has “an equal chance to

\begin{thebibliography}{99}
\footnotesize


64 Hill (1980, 92 f.) discusses the question of whether the worth of several beings possessing dignity can be compared with each other although it cannot be compared with things possessing no dignity. He discriminates the following interpretations (laid out in questions): “Is his [Kant’s] view, for example, that there are two scales of value, price and dignity, such that things can be ranked comparatively on each scale even though nothing on the scale of dignity can be overweighed by any amount of value on the scale of price?” And: “[I]s Kant’s view that dignity is something that cannot be quantified, so that it does not make sense to say that dignity of humanity in one person can fairly and reasonably be exchanged for the sake of a greater amount of dignity elsewhere?” As I see it (but here cannot really argue for), Kant’s ground-thesis (cf. Schmidt/Schönecker 2018) implies that there can be no scale of dignity.

65 For the question of whether counting numbers can be reconciled with deontological positions, cf. Hevelke/Nida-Rümelin (2015); for the question of whether consequentialism really implies that one ought to kill the one, cf. Howard (2021).

\end{thebibliography}
survive” (Taurek 1977, 303). At first glance, this position might seem outlandish (or counterintuitive, if you will), but on second thought, it is worth considering. Although it is not undisputed what Kant’s response to Taurek cases would be, I submit it would be the following: Given that a group of ten beings that possess dignity is by no means of greater worth than a group of five such beings because there is no scale of dignity, there is no moral reason to save the larger group rather than the smaller one. With Kant, it is certainly permissible to save the ten, but it is not obligatory: it is also permissible to save the five. On Kantian grounds, therefore, Foot’s rule 2 cannot be correct; it is not true that numbers count in situations involving the same kind of duties (helping ten vs. helping five). A fortiori, we have good reason to argue that in a trolley situation (which, again, is different from a Taurek case) the numbers are completely irrelevant; yet not only, nota bene, because of the equal worth of five vs. ten persons, but also because the set of duties at stake is different (helping vs. killing). Even if Kant did accept the obligation to save the ten and not the five when it comes to Taurek cases, this would not imply the obligation to save the five and kill the one in the trolley situation.

67 In recent literature, different kinds of Taurekian positions are discussed. Dufner/Schöne-Seifert (2019, 27) distinguish between a strong version (as explained above) and a weighted version (according to which the ten people should have a chance of rescue twice as big as that of the five because the group is twice the size). Additionally, they discuss the Tie Breaker position (stemming from Scanlon) and their own. However, only the first, strong version is the Taurekian version in a strict sense. For the discussion concerning Taurek and the question of whether saving the greater number is required, cf., amongst others, Timmermann (2004) and Meyer (2006).


69 Timmermann (2004, 110) shares the opinion that from a Kantian perspective numbers do not count. However, he suggests another quite enlightening method to determine whom to save which he calls individualist lottery (he describes a situation in which you can save A on one island or B and C on another island): “To give the claims of A, B and C equal weight, a coin will not do. We need a wheel of fortune with three sectors, each of which bears the name of one islander. The person whose sector comes up is saved. If this person is A, both B and C perish. Having reached the island, the rescuer then incurs an obligation to save C. Similarly, if C wins B is also saved. We neither count, nor aggregate, nor quantify; nor do we arbitrarily assign roles to individual islander.” – On a different note, stressing the impossibility of counting numbers when it comes to beings that possess dignity, it might seem impossible to say that the death of ten (either through murder or a natural disaster, for instance) is worse than the death of a single person (cf. Dufner/Schöne-Seifert 2019, 26). I am not sure whether this is true (and I am not sure if it posed a problem if it were true), but I cannot get into this discussion at this point.
6 Conclusion. Kant and the Programmer

How should autonomous vehicles be programmed with regard to trolley-like situations against the background of Kantian ethics? As we have seen, there is no normative difference between the bystander and the driver of the trolley because by not intervening neither of them kills the five in a morally relevant way. Instead, whoever turns the trolley must choose between the wide duty to help (the five) and the narrow duty not to kill the one. Since the narrow duty trumps the wide duty, and because the general wide duty to help others (that allows for a certain amount of latitude) is different from a specific obligation to help the five, it is not permissible to flip the switch – no conflict of duties is involved. Hence, it is also not permissible for the Kantian programmer of an autonomous vehicle to make his algorithm steer the car around a group of people if doing so results in hitting a smaller group or a single person. I do not claim that indeed it is morally obligatory not to flip the switch – my only aim was to show what a Kantian position looks like.

The Kantian solution to the trolley problem differs considerably from Foot’s. Although both Foot and Kant take narrow duties to trump wide duties (Foot’s rule 1), there is no Kantian way to interpret the situation as one only concerning negative duties; whoever has to pull the switch is not to choose between killing the five and killing the one. Furthermore, Kant presumably would not have accepted Foot’s rule 2, i. e., the rule that the number of victims is to be minimized, neither in a case involving the same kind of duties (Taurek cases) nor in a situation in which one would have to kill one person to save others (trolley). Kant’s concept of dignity as an absolute value forbids weighing one life against another, regardless of the numbers.

Now let us take a step back. We moved from highly abstract trolley cases and thus from philosophical thought experiments to ethical puzzles in quite mundane situations concerning autonomous driving. One might wonder whether this is a good idea at all. There are at least two possible objections here. First, it might be argued that the whole endeavor of trolleyology is misguided, or at least misguided in the way it is actually done. Philosophers take their intuitions about whether it is permissible to flip the switch as their irrefutable and so in some sense axiomatic starting points of argumentation, searching for theories why it is permissible – according to their intuitions – to pull the switch in one case but not in another. Whilst doing so, they do not question these intuitions – which they

Foot (1978 [1967], 27) actually speaks of negative and positive duties but, at least roughly speaking, she means narrow and wide duties.
should, according to the first critique. For there are reasonable doubts concerning the epistemic reliability of our intuitions, both in general and with regard to trolley cases in particular. Thus, Joshua Greene (2016, 176; Greene et al. 2009) argues that our intuitions concerning trolley cases might correlate with factors which are not morally relevant at all but whose significance for us is rather due to other circumstances. For example, people seem to be more critical about actions that harm other people when carried out through personal strength (pushing the large man down the bridge with my own hands as opposed to pushing a button which initiates a causal reaction leading to the fall of the large man). This could explain why people tend to say it is permissible to flip the switch in the loop case but not in the large man case. Also, our intuitions might at least partly depend, as Di Nucci (2014) argues, on the order of the cases presented. While most participants in his study judge it to be permissible to pull the switch in the original bystander’s case when confronted first or only with this case, judgements change a great deal when people are presented first with Thomson’s three options case (in which the bystander has a choice between the death of the five, the death of the one, and his own death) and only afterwards with the bystander’s original case; then the majority indeed opts not to flip the switch in the bystander’s case. All this might lead us to think, as Peter Singer (2005) argues, that we should not use arbitrary intuitions to refute ethical theories but use ethical theories to refute our intuitions.

Secondly, from an application-oriented point of view, there are serious doubts about the reasonableness of comparing hypothetical trolley situations with situations that autonomous vehicles might get into. For several reasons, trolley cases, it is argued, are unlikely to be of any help in real life. I concentrate here on three aspects. (i) Trolley cases do not occur in reality. A rather specific variant of this critique was put forward by Johannes Himmelreich (2018, 674): For technical reasons, he says, “there seems to be a tension between having a meaningful choice and a collision being unavoidable”, which is to say that it is unlikely that one can steer a trolley if the brakes have completely failed. (ii) In reality, we cannot be sure whether endangered persons really do have no chance of survival at all. Rather, there are only different probabilities for different situations (we usually cannot be sure whether a person hit by a car will certainly die, nor can we be sure

72 According to Singer (2005, 350), such a revisionist approach to intuitions leads us to the conclusion that throwing the large man down the bridge is indeed the right thing to do.
73 For the following aspects of critique and further aspects, cf. Goodall (2016), Nyholm/Smids (2016), Himmelreich (2018), and Keeling (2020, though he ultimately defends the advantages of discussing trolley problems for autonomous driving).
whether the car will hit the person at all). (iii) Trolley problems get much more complicated when more details are known, for instance, the age of the persons involved, their gender, their health conditions, their role in society, their criminal record, their social and familial status, or their role or responsibility in the given situation. – Since there are these differences between trolley and real-world situations, critics say, it does not make any sense to talk about trolleys when what one really wants to talk about are real-world situations.

How can these charges be addressed? To begin with the objection of applicability: It is crucial to see that from a Kantian point of view, quite a number of those real-life aspects mentioned above are irrelevant. Human beings possess dignity no matter how old or healthy they are, for instance, and thus the relevance of the objection is dependent on the ethical outlook one has. Furthermore, although it is certainly true that real-life situations are more complicated, thinking about trolley cases makes us see essential aspects of the ethical problem to be solved and this asset is not undermined by the fact that in reality more features of the situation have to be taken into account.74 Also, it is of course true that to consider risks and probabilities in the trolley scenario would make the analysis much more complicated.75 Nevertheless, solving the trolley problem in a merely theoretical manner is indeed a sensible thing to do since exploring the theoretical framework of a practical problem does help to solve the practical problem. This is true even if trolley situations in the real world are somewhat more rocky.

A final word on intuitions. The question of what intuitions are and what epistemic role they play in the game of justification has been disputed ever since the very idea of something like intuition or maybe self-evidence came up; more recently, experimental philosophy has shed new doubts on the reliability of intuitions. What we can learn indeed from a Kantian analysis is that although intuitions cannot (and should not) be banned completely from practical philosophy, we should not rely on them blindly. Rather, we should be ready to dismiss some of them, and this is exactly what a Kantian analysis of the trolley problem brings home. Although it certainly appears to many intuitively clear that flipping the switch in the original trolley cases is permissible, it came to light that on Kantian grounds flipping the switch really is not permissible – or at least Kant’s approach is capable of shaking our intuition that it is. In many respects, if one wants to sustain Kantian theory, it would be more sensible to give up the intuition that it

75 From a Kantian perspective, such an analysis is provided by Bjorndahl/London/Zollmann (2017).
is permissible to flip the switch and try to make our judgements fit our theories rather than to try desperately to make Kantian theory fit our intuitions. This is not to say, however, that Kant’s ethics does not rely on some kind of intuition itself.

References

I use the following abbreviations for Kant’s writings:

GMS  
*Groundwork of the Metaphysics of Morals*

KpV  
*Critique of Practical Reason*

MS  
*The Metaphysics of Morals*

NG  
*Attempt to Introduce the Concept of Negative Magnitudes into Philosophy* (AA 2)

PPH  
*Practical Philosophy* Herder

TP  
*On the Common Saying: That may be Correct in Theory, But it is of no Use in Practice*

VRML  
*On a Supposed Right to Lie from Philanthropy*


PPH, my own translation.


